

Annotations *Rhapsodie* pour le *Trameur*

21/03/2015 21:53:35

Serge Fleury

(document de travail)

Références

- **Le Trameur, manuel d'utilisation**

<http://www.tal.univ-paris3.fr/trameur/leMetierLexicometrique.pdf>

Dans cette documentation, la partie « *Relations de dépendance entre les items de Trame (via leurs annotations)* » présente l'état des développements actuellement disponibles dans le *Trameur* pour l'exploitation des relations de dépendance.

- **Le Trameur. Propositions de description et d'implémentation des objets textométriques**

<http://www.tal.univ-paris3.fr/trameur/trameur-propositions-definitions-objets-textometriques.pdf>

Ce document met au jour une description des objets textométriques et les méthodes mises en œuvre dans le *Trameur* pour travailler sur et avec ces objets dans une perspective textométrique. On y détaille aussi les opérations permises sur une *base textométrique* : format des données textuelles, modification dynamique de la *Trame*, correction ou ajout d'annotation etc.

- **Description du format tabulaire du TreeBank Rhapsodie** (Ilaine Wang, Rachel Bawden). Version : morpho-syntaxe, micro-syntaxe, macro-syntaxe, prosodie) March 20, 2015

- **Annotations Rhapsodie pour le Trameur**

<http://www.tal.univ-paris3.fr/trameur/bases/rhapsodie2trameur.pdf>

<http://www.tal.univ-paris3.fr/trameur/bases/rhapsodie2trameur-v4.pdf>

<http://www.tal.univ-paris3.fr/trameur/bases/rhapsodie2trameur-v6.pdf>

Ces documents présentent les différentes étapes du processus de transcodage des annotations Rhapsodie pour construire une base Textométrique ; présentation des différents processus de traitements des annotations de dépendance.

Le contenu du document actuel complète et enrichit la présentation faite dans les 2 références précédentes.

1. Préambule

Ce document commence par décrire le processus de transcodage des données issues du projet *Rhapsodie* (<http://projet-rhapsodie.fr/>) sous la forme d'une base textométrique importable dans le *Trameur*. Il présente ensuite différentes fonctionnalités mises en œuvre pour traiter ce type de données via le *Trameur*.

2. Données Rhapsodie

Les données traitées sont disponibles sur la page du projet *Rhapsodie*.

Par exemple ici :

- Téléchargement des fichiers de codage microsyntactique version bêta 10/13 ([zip](#))
- Tutoriel codage microsyntactique ([pdf](#))

Le fichier complet d'annotations a l'allure suivante (lecture partielle ici dans un tableur) :

Text_id	Tree_ID	Token_ID	Token	Speaker	Wordform	Wordform	Lemma	POS	Morph	Trace	Person	Number	Gender	ID_dep	Type_dep	ID_pplan	Type_pplan	ID_junc	Type_junc	ID_pana	Type_pana	ID_junc_inherited	Type_junc_inherited	ID_junc_inherited	Type_junc_inherited	Layer	Associed_andus	Intro_UI	Period	Period_tone	Package	Package_tone						
D0002	1	1	euh	SL1	B	euh	euh	I						0	root	0	root																	filled-pause	_			
D0002	1	2																																				
D0002	1	3	ils	SL1	B	ils	il	CI			3	pl	mascul	5	sub	5	sub																		included	LIL1		
D0002	1	4																																				
D0002	1	5	avaient	SL1	B	avaient	avoir	V	indicative	imperfect	3	pl		0	root	0	root																			included	LIL1	
D0002	1	6																																				
D0002	1	7	&		B	&	&	N					mascul/fem	5	obj	5	obj																					
D0002	2	1	j	SL1	B	j'	je	CI			1	sg		3	sub	3	sub																			included	LIL1	
D0002	2	2	'	SL1	I																																included	LIL1
D0002	2	3	avais	SL1	B	avais	avoir	V	indicative	imperfect	1	sg		0	root	0	root																				included	LIL1

Ce fichier tabulé est composé de 59 colonnes.

Ces données sont constitués par un certain nombre de textes (l'identifiant du texte (Text_id) est visible dans la première colonne), chacun d'eux est segmenté en « unité illocutoire » (UI, seconde colonne), chacune d'elle est segmentée en token (troisième colonne), chacun d'eux est annoté (les autres colonnes).

Les données issues de Rhapsodie sont décrites dans le tableau suivant :

source : Description du format tabulaire du TreeBank *Rhapsodie* (Ilaine Wang, Rachel Bawden)

1	Text ID	le nom du texte (D0001, M2006 etc.)
2	Tree ID	le numéro de l'arbre dans le texte
3	Token ID	le numéro du token dans l'arbre
4	Token	la forme du token. Des lexèmes composés de plusieurs mots orthographiques ont été segmentés en tokens individuels. Un token est donc un segment de la transcription compris entre deux blancs ou un blanc et un signe de ponctuation. Tous les caractères qui ne sont pas des lettres (les espaces, les tirets et les apostrophes) sont considérés comme des tokens individuels aussi.
5	Speaker	l'identifiant du locuteur. En cas de chevauchement, on peut avoir plusieurs locuteurs (annotés alors \L1-\L3 par exemple).
6	Word span	la position du token dans le mot forme. La valeur est soit B (begin) pour le premier token d'un mot, soit I (inner) pour les tokens qui en sont pas les premiers tokens du mot.
7	Word form	le mot-forme auquel appartient le token. Dans le cas d'un mot forme comprenant plusieurs tokens, le mot-forme est uniquement marqué pour le premier token.
8	Lemma	le lemme du lexème auquel appartient le token. Dans le cas où il y a plusieurs tokens qui font partie du même lexème, le lemme n'est pas répété : le lemme est écrit dans cette colonne pour la ligne correspondant au premier token du lexème
9	POS	la catégorie morpheo-syntaxique du mot auquel appartient le token parmi N, V, Adj, Adv, I, P, D, Cl, Pro, CS, Qu.
10	Mood	le mode pour les verbes parmi 'indicative', 'subjunctive', 'imperative', 'past_participle' et 'present_participle'. Dans le cas où la forme est ambiguë, les deux possibilités de mode sont indiquées (ex : indicative/subjunctive).
11	Tense	le temps grammatical du verbe parmi 'present', 'future', 'conditional', 'imperfect' et 'perfect'. Le temps est marqué uniquement pour les verbes qui ont pour mode 'indicative'.
12	Person	la personne grammaticale pour les verbes et les pronoms personnels, ('1', '2' ou '3'). En cas d'ambiguïté, les personnes possibles sont toutes écrites séparées par des barres obliques (ex : '1/2/3').
13	Number	le nombre grammatical ('sg' ou 'pl' ou 'sg/pl' en cas d'ambiguïté) pour les verbes conjugués, les noms, les adjectifs, les pronoms et certains mots qu- (quel, quels, laquelle etc.).
14	Gender	le genre grammatical ('masc', 'fem' ou 'masc/fem' en cas d'ambiguïté) pour les noms, les adjectifs, les participes passés et certains mots qu- (quel, quels, laquelle etc.).
15	ID dep	le numéro du gouverneur par dépendance. Le numéro du gouverneur correspond à la colonne Token ID. Dans le cas où un gouverneur est constitué de plusieurs tokens, c'est le Token ID du premier token qui est pris comme numéro de gouverneur. Ce principe tient aussi pour les autres types de liens de dépendance.
16	Type dep	le type de lien de dépendance correspondant à ID dep.
17	ID plain	le numéro du gouverneur par dépendance 'primitive'.
18	Type plain	le type de lien de dépendance (primitif), correspondant à ID plain (correspondant aux liens pred, root, sub, dep, obj, obl, ad). N.B. Il ne peut y avoir qu'un seul type de dépendance primitive et un seul gouverneur primitif par token.
19	ID para	le numéro du gouverneur par lien paradigmatique.
20	Type para	le type de lien paradigmatique (parmi les types para_disfl, para_coord, para_intens, para_dform, para_reform, para_hyper, para_negot) N.B. Il ne peut y avoir qu'un seul type de dépendance paradigmatique et un seul gouverneur

		paradigmatique par token.
21	Gov inher	le numéro du gouverneur par lien hérité.
22	Type inher	le type de lien hérité (parmi <code>pred_inherited</code> , <code>root_inherited</code> , <code>sub_inherited</code> , <code>dep_inherited</code> , <code>obj_inherited</code> , <code>obl_inherited</code> , <code>ad_inherited</code>). N.B. Il ne peut y avoir qu'un seul type de dépendance par token, mais il peut y avoir plusieurs gouverneurs par dépendance héritée. Dans ce cas, les numéros des gouverneurs sont séparés par une virgule. Ex :
23	Gov junc	le numéro du gouverneur par lien 'junc' (de jonction)
24	Type junc	le type de lien junc - il n'y en a qu'un seul, donc ceci correspond toujours à junc. Cette colonne est ici pour l'uniformité du tableau.
25	Gov junc-inher	le numéro du gouverneur par lien 'junc inherited' (de jonction héritée)
26	Type junc-inher	le type de lien junc inherited - il n'y en a qu'un seul, donc ceci correspond toujours à <code>junc_inherited</code> . Cette colonne est ici pour l'uniformité du tableau.
27	Layer	l'appartenance à un entassement. Dans cette annotation, les différents niveaux d'entassement sont écrasés.
28	IU	l'appartenance à une unité illocutoire (UI, ou Illocutionary Unit en anglais). Cette information est bien fournie au format BILOU, mais on note que pour cette unité, la valeur O n'est pas utilisée du fait que tous les tokens font nécessairement partie d'une UI.
29	Nucleus	l'appartenance à un noyau.
30	Prenucleus	l'appartenance à un pré-noyau.
31	Gov prenucleus	l'appartenance à un pré-noyau régi, c'est-à-dire qui fait partie de la même unité rectionnelle que l'UI à laquelle il est rattaché.
32	Innucleus	l'appartenance à un in-noyau.
33	Gov innucleus	l'appartenance à un in-noyau régi.
34	Postnucleus	l'appartenance à un post-noyau.
35	Gov postnucleus	l'appartenance à un post-noyau régi.
36	IU parenthesis	l'appartenance à une UI parenthétique.
37	IU graft	l'appartenance à une greffe, qui correspond généralement à un discours rapporté ou bien à des séquences qui contribuent à remplir la position syntaxique d'un élément recteur.
38	IU embedded	l'appartenance à une unité enchâssée qui ne correspond pas à une UI (alors que c'est le cas de la greffe).
39	Associative nucleus	l'appartenance à une liste fermée de marqueurs discursifs établie pour le projet.
40	Intro IU	l'appartenance à une unité qui permet d'introduire une UI.
41	Period	l'appartenance à une période intonative. Cette information est donnée au format BILOU (décrit dans l'introduction de la section précédente) sans le O comme c'était le cas pour l'UI (28), puisque, de la même manière, tout token fait partie d'une période intonative. On y trouve toutefois un détail supplémentaire : en plus de marquer la position du token dans la période, chaque valeur peut également indiquer s'il s'agit d'une période tronquée à gauche (-B, -I, -L, -U), à droite (B,-I,-L,-U,-), des deux côtés (-B,-I,-L,-U-).
42	Period tone	le profil tonal (ou contour intonatif) de la période intonative dans laquelle se trouve le token.
43	Package	le type de paquet intonatif dans lequel le token se trouve parmi <code>filled-dis</code> , <code>filled-pause</code> , <code>included</code> , <code>lone</code> , <code>lone-dis-strong</code> , <code>motherless</code> , <code>motherless-dis-weak</code> , <code>silent-pause</code> , ou <code>tail</code> .
44	Package tone	le profil tonal (ou contour intonatif) du paquet intonatif dans lequel se trouve le

		token.
45	Group	le type de groupe rythmique dans lequel le token se trouve parmi dis-strong, dis-weak, filled-dis, filled-pause, silent-pause, strong, tail, ou weak.
46	Group tone	le profil tonal (ou contour intonatif) du groupe rythmique dans lequel se trouve le token.
47	Foot	le dernier pied métrique du token. à l'instar du groupe rythmique, il peut être annoté dis-strong, dis-weak, filled-dis, filled-pause, silent-pause, strong, tail, ou weak.
48	Foot tone	le profil tonal (ou contour intonatif) du dernier pied métrique du token.
49	Syllable tone	le profil tonal (ou contour intonatif) de la dernière syllabe du token.
50	Prominence initial	le degré de proéminence de la première syllabe du token. Une proéminence peut être annotée W pour Weak ou S pour Strong. Dans le cas où cette syllabe n'est pas proéminente, elle peut avoir comme valeur 0, _ (pause) ou encore % (syllabe inaudible ou non transcrite en raison d'un chevauchement).
51	Prominence final	le degré de proéminence de la dernière syllabe du token. Elle peut avoir les mêmes valeurs que pour la proéminence initiale.
52	Hesitation	marquée H pour la particule \"eh\" ou bien pour une syllabe hésitante, _ (pause) ou bien % (mot inaudible ou non transcrit en raison d'un chevauchement).
53	Pause length	la durée de la pause (en s), indiquée au niveau du token qui la précède. Autrement dit, si cette donnée est vide pour un token donné, alors il n'est pas suivi d'une pause. à noter que dans le cas précis d'un chevauchement et si les paroles du locuteur principal sont transcrites en premier, la longueur de la pause est indiquée sur le dernier mot, à la _n du chevauchement mais on a ajouté le signe dièse sur la ligne du mot après lequel on observe réellement une pause. Ce choix a été fait afin de préserver un alignement pertinent avec les différentes unités d'analyse, notamment la période.
54	Tmin	le temps de début de chaque token (ou le cas échéant, du chevauchement) au sein de son échantillon.
55	Tmax	le temps de fin de chaque token (ou le cas échéant, du chevauchement) au sein de son échantillon.
56	Syllable length	la durée de la dernière syllabe du token (en ms).
57	Syllable length avg	la durée moyenne de cette syllabe (en ms).
58	Pitch	la hauteur (en demi-tons) de cette syllabe.
59	Pitch avg	la hauteur moyenne (en demi-tons), calculée sur un empan de quelques syllabes précédant et suivant la dernière syllabe du token.

Les annotations sont réutilisées telles quelles par le processus de transcodage vers *le Trameur* sauf pour les annotations de dépendance (16 à 26) qui sont réutilisées 2 à 2 (*cf* jeu de couleur) pour construire respectivement une seule annotation (de type relation) qui est réécrite par exemple sous la forme : `Type_rection(Gov-rection)`.

CIBLE est une valeur numérique pointant vers une position de la *Trame*

Chaque item de la base textométrique est associé *in fine* à 49 annotations décrites dans le tableau suivant :

n°Annotation	Annotation trameur	Contenu Rhapsodie
1	Forme	Token
2	Lemme	Lemma
3	Cat	POS
4	a-00004	Word_span
5	a-00005	Mood
6	a-00006	Tense
7	a-00007	Person
8	a-00008	Number
9	a-00009	Gender
10	a-00010	Type_dep(ID_dep)
11	a-00011	Type_plain(ID_plain)
12	a-00012	Type_junc(ID_junc)
13	a-00013	Type_para(ID_para)
14	a-00014	Type_inherited(ID_inherited)
15	a-00015	Type_junc_inherited(ID_junc_inherited)
16	a-00016	Layer
17	a-00017	IU
18	a-00018	Nucleus
19	a-00019	Prenucleus
20	a-00020	Gov_nucleus
21	a-00021	Innucleus
22	a-00022	Gov_innucleus
23	a-00023	Postnucleus
24	a-00024	Gov_postnucleus
25	a-00025	IU_parenthesis
26	a-00026	IU_graft
27	a-00027	IU_embedded
28	a-00028	Associated_nucleus
29	a-00029	Intro_IU
30	a-00030	Prominence_final
31	a-00031	Prominence_initial
32	a-00032	Hesitation
33	a-00033	Pitch_avg
34	a-00034	Pitch
35	a-00035	Syllabe_tone
36	a-00036	Syllabe_length

37	a-00037	Syllabe_length_avg
38	a-00038	Speaker
39	a-00039	Pause_length
40	a-00040	Period
41	a-00041	Period_tone
42	a-00042	Package
43	a-00043	Package_tone
44	a-00044	Group
45	a-00045	Group_tone
46	a-00046	Foot
47	a-00047	Foot_tone
48	a-00049	Tmin
49	a-00049	Tmax

La figure suivante donne à voir, dans *le Trameur*, les annotations disponibles sur un item en contexte (le mot « lycée ») :

The screenshot displays the 'Le Trameur - Le Métier Lexicométrique @CLA2T-P3 V.12.44' interface. The main window is divided into several panes:

- Left Panel:** Contains controls for 'Chargement de la Carte des sections', 'Délimiteur de sections' (set to '\$'), and 'Parties' (listing PARTIE, GENRE, SUBGENRE). It also includes a 'Recherche Forme sur la carte' section and 'Spécificités sur Sections'.
- Center Panel:** Shows a grid of sections with labels like D0002, M0009, M0016, D2011, M0022, M2005, M0012, D2004, and M0004. Below this is a text area showing 'le lycée Voltaire est un bon lycée'.
- Right Panel (Annotations):** A list of linguistic annotations for the word 'lycée', including:
 - Position:<532>
 - Forme:<lycée>|Freq:19
 - Lemme:<lycée>|Freq:22
 - Cat:<N>|Freq:6311
 - a-00004:|Freq:34491
 - a-00005:<->|Freq:79363
 - a-00006:<->|Freq:81009
 - a-00007:<->|Freq:76646
 - a-00008:<sg>|Freq:17714
 - a-00009:<masc>|Freq:9217
 - a-00010:<PRED(526)>|Freq:1
 - a-00011:<PRED(526)>|Freq:1
 - a-00012:<->|Freq:37744
 - a-00013:<->|Freq:37265
 - a-00014:<->|Freq:37523
 - a-00015:<->|Freq:38411
 - a-00016:<O>|Freq:25855
 - a-00017:<L>|Freq:2864
 - a-00018:<L>|Freq:2871
 - a-00019:<O>|Freq:35163
 - a-00020:<O>|Freq:36100
 - a-00021:<O>|Freq:37537
 - a-00022:<O>|Freq:37890
 - a-00023:<O>|Freq:37278
 - a-00024:<O>|Freq:37682
 - a-00025:<O>|Freq:36644
 - a-00026:<O>|Freq:36789
 - a-00027:<O>|Freq:36531
 - a-00028:<O>|Freq:35917
 - a-00029:<O>|Freq:36976
 - a-00030:|Freq:11601
 - a-00031:<O>|Freq:22703
 - a-00032:<->|Freq:81018
 - a-00033:<86.13582490021034>|Freq:1
 - a-00034:<85.84795653167001>|Freq:1
 - a-00035:<L1>|Freq:351
 - a-00036:<772.9000000000007>|Freq:1
 - a-00037:<117.67724999999984>|Freq:1
 - a-00038:<L1>|Freq:24688
 - a-00039:<->|Freq:82117
 - a-00040:<I>|Freq:32455
 - a-00041:<Lm>|Freq:257
 - a-00042:<included>|Freq:16798
 - a-00043:<lll1>|Freq:22
 - a-00044:<45>|Freq:38423
 - a-00045:<ll>|Freq:2226
 - a-00046:|Freq:18545
 - a-00047:<48>|Freq:38423
 - a-00048:<66.1479>|Freq:1
 - a-00049:<66.9808>|Freq:1
- Bottom Panel:** Shows 'Annotations : 1 2 3' and a status bar with keyboard shortcuts.

3.3 Le Cadre textométrique

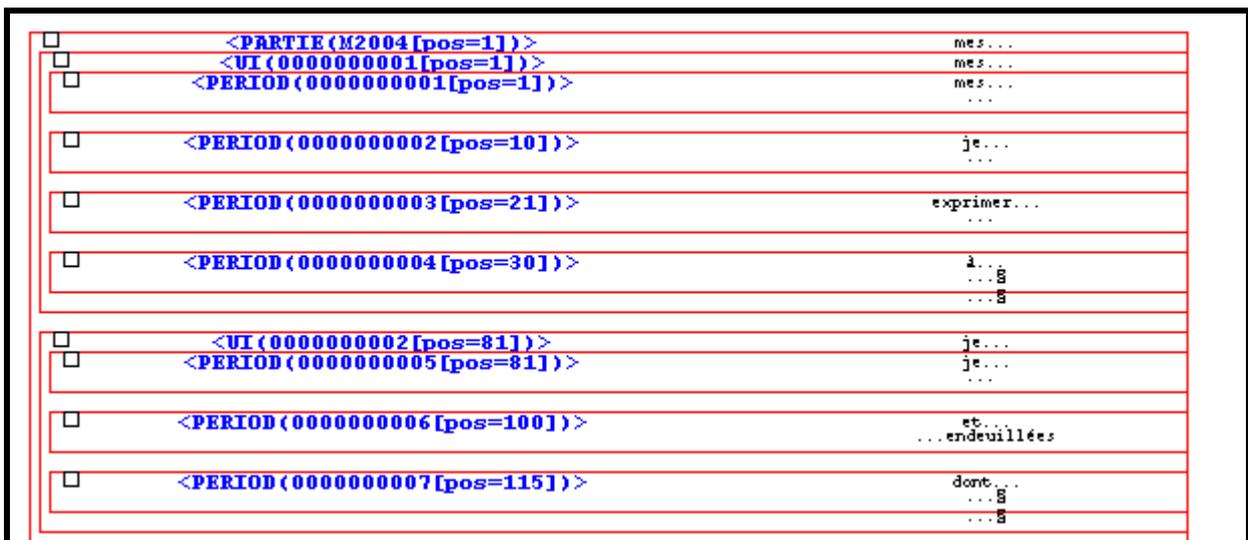
Les différents échantillons initiaux de *Rhapsodie* sont considérés comme autant de parties différentes : la base finale est donc une partition de textes (*Cadre*), chaque partie contient les zones textuelles associées à l'identifiant initial de l'échantillon.

Deux autres partitions permettent de contraster certaines zones textuelles du corpus : les UIs et les périodes.

Découpage des échantillons

- **PARTIE** : partition des échantillons
- **UI** : partition des UI
- **PERIOD** : partition des périodes

La figure suivante illustre l'organisation des différentes parties et leur structuration logique :



Un échantillon est donc composé d'un ensemble d'UI, chaque UI est composé d'un ensemble de période.

Catégorisation des échantillons

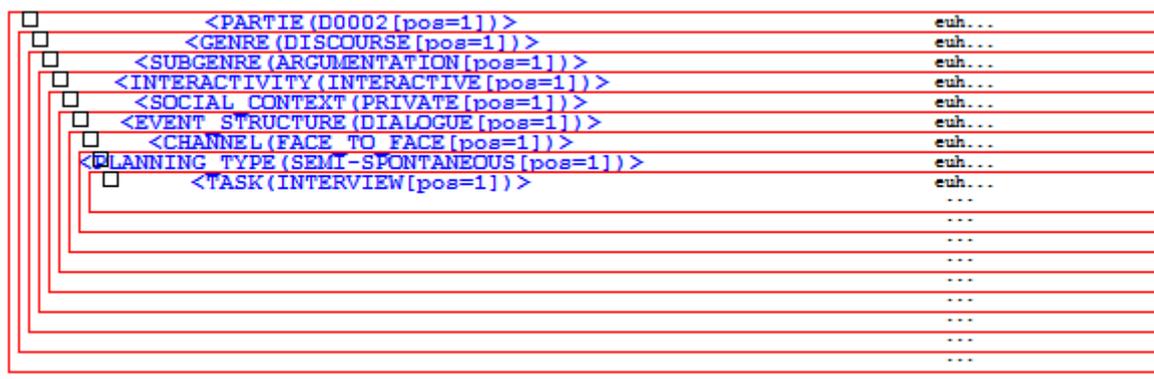
Chaque échantillon (*cf* Text_id vu précédemment) est catégorisé suivant différents points de vue : genre, type de conversation etc. Les différentes variables permettant de catégoriser les échantillons sont regroupées dans le tableau suivant :

Text_id	Subgenre	Interactivity	Social_context	Event_structure	Channel	Planning_type	Task
D0001	Argumentation	interactive	Private	Dialogue	Face_to_Face	semi-spontaneous	interview
D0002	Argumentation	interactive	Private	Dialogue	Face_to_Face	semi-spontaneous	interview
D0003	Description	semi-interactive	Private	Dialogue	Face_to_Face	spontaneous	interview
D0004	Description	interactive	Private	Dialogue	Face_to_Face	semi-spontaneous	interview
D0005	Description	semi-interactive	Private	Dialogue	Face_to_Face	spontaneous	interview
D0006	Argumentation	interactive	Private	Dialogue	Face_to_Face	semi-spontaneous	interview
D0007	Procedural	non-interactive	Private	Dialogue	Face_to_Face	spontaneous	travel-planning
D0008	Procedural	non-interactive	Private	Dialogue	Face_to_Face	spontaneous	travel-planning
D0009	Description	interactive	Private	Dialogue	Face_to_Face	spontaneous	Unknown
D0017	Procedural	non-interactive	Private	Dialogue	Face_to_Face	spontaneous	travel-planning
D0020	Procedural	non-interactive	Private	Dialogue	Face_to_Face	spontaneous	travel-planning
D1001	Description	semi-interactive	Private	Dialogue	Face_to_Face	planned	Unknown
D1002	Description	semi-interactive	Private	Dialogue	Face_to_Face	spontaneous	lesson
D1003	Description	semi-interactive	Private	Dialogue	Face_to_Face	spontaneous	life_story
D2001	Narrative	interactive	Public	Dialogue	Broadcasting	semi-spontaneous	interview
D2002	Argumentation	interactive	Public	Dialogue	Broadcasting	planned	interview
D2003	Description	semi-interactive	Public	Dialogue	Broadcasting	spontaneous	sportscasting
D2004	Narrative	semi-interactive	Public	Dialogue	Broadcasting	spontaneous	life_story
D2005	Narrative	semi-interactive	Public	Dialogue	Broadcasting	planned	interview
D2006	Oratory	semi-interactive	Public	Dialogue	Face_to_Face	planned	Political_speech
D2007	Argumentation	interactive	Public	Dialogue	Broadcasting	spontaneous	Conversation
D2008	Argumentation	interactive	Public	Dialogue	Broadcasting	spontaneous	interview
D2009	Argumentation	semi-interactive	Public	Dialogue	Broadcasting	planned	interview
D2010	Narrative	interactive	Public	Dialogue	Broadcasting	spontaneous	interview
D2011	Argumentation	semi-interactive	Public	Dialogue	Broadcasting	planned	Advertising
D2012	Description	interactive	Public	Dialogue	Broadcasting	spontaneous	interview
D2013	Argumentation	interactive	Public	Dialogue	Broadcasting	spontaneous	interview
M0001	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0002	Description	semi-interactive	Private	Monologue	Face_to_Face	spontaneous	movie_scene_description
M0003	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0004	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0005	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0006	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0007	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0008	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0009	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning

M0010	Procedural	non-interactive	Private	Dialogue	Face_to_Face	spontaneous	travel-planning
M0011	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0012	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0013	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0014	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0015	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0016	Procedural	non-interactive	Private	Monologue	Face_to_Face	spontaneous	travel-planning
M0018	Description	semi-interactive	Private	Monologue	Face_to_Face	spontaneous	movie_scene_d escription
M0019	Description	semi-interactive	Private	Monologue	Face_to_Face	spontaneous	movie_scene_d escription
M0021	Description	semi-interactive	Private	Monologue	Face_to_Face	spontaneous	movie_scene_d escription
M0022	Description	semi-interactive	Private	Monologue	Face_to_Face	spontaneous	movie_scene_d escription
M0023	Description	semi-interactive	Private	Monologue	Face_to_Face	spontaneous	movie_scene_d escription
M0024	Description	semi-interactive	Private	Monologue	Face_to_Face	spontaneous	movie_scene_d escription
M1001	Description	semi-interactive	Private	Monologue	Face_to_Face	spontaneous	life_story
M1003	Description	semi-interactive	Private	Monologue	Face_to_Face	spontaneous	professional_pr oject_descriptio n
M2001	Oratory	non-interactive	Public	Monologue	Broadcasting	planned	political_speech
M2002	Description	non-interactive	Public	Monologue	Broadcasting	planned	Talk
M2003	Oratory	non-interactive	Public	Monologue	Broadcasting	planned	sermon
M2004	Oratory	non-interactive	Public	Monologue	Broadcasting	planned	political_speech
M2005	Description	non-interactive	Public	Monologue	Broadcasting	planned	press_review
M2006	Description	non-interactive	Public	Monologue	Broadcasting	planned	info-kiosk

Les différentes valeurs associées à chaque échantillon permettent donc de re-partitionner le corpus suivant les sept variables associées à chacune des colonnes du tableau précédent : Subgenre, Interactivity, Social_context, Event_structure, Channel, Planning_type, Task

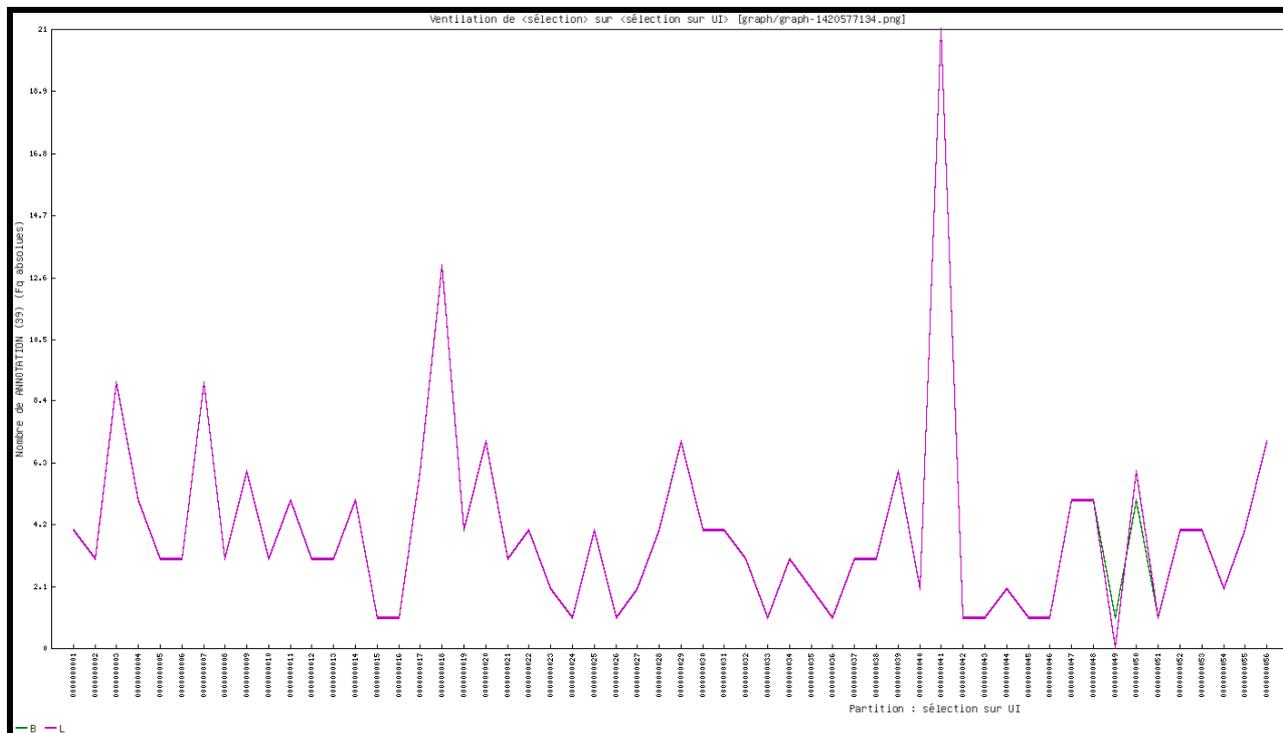
Au final chaque échantillon peut être vu suivant les différentes métadonnées qui lui sont associées :



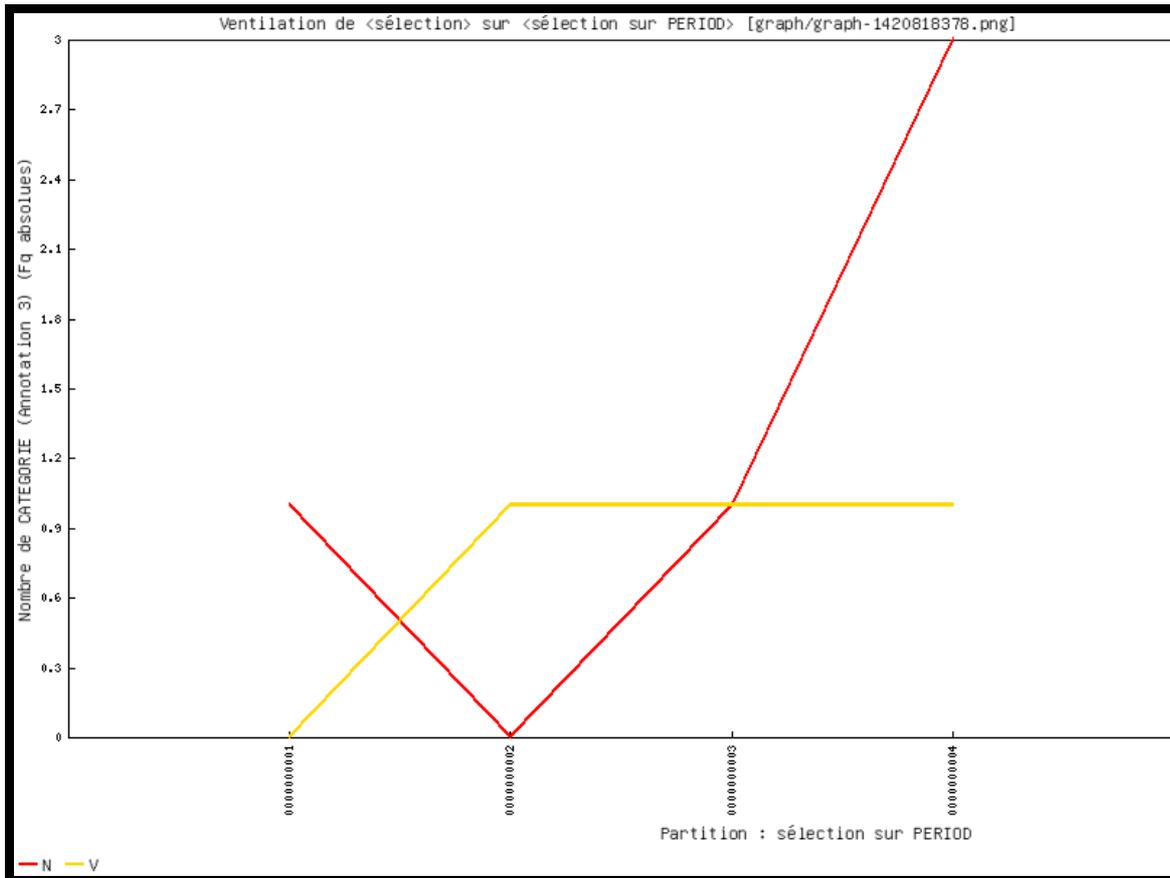
Les différentes unités du corpus pourront donc visualisées ou distribuées suivant ces différents points de vue (*cf infra*).

4. Ventilation : distribution d'unités

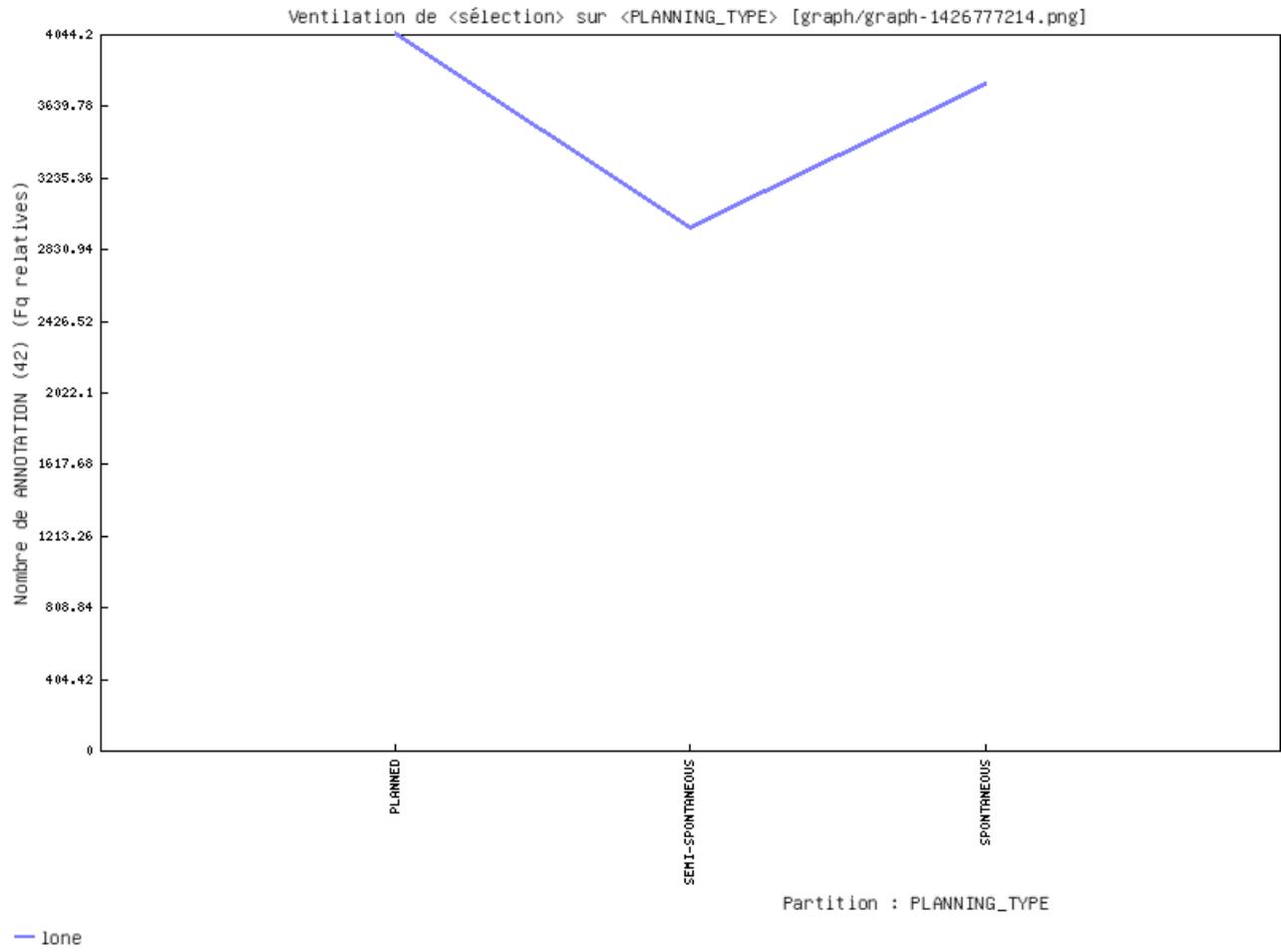
L'introduction de deux nouveaux systèmes de parties permet désormais de construire des graphiques de ventilation sur ces parties. La figure suivante met au jour par exemple la ventilation des items B et L (annotation n°40 : Period) sur la partition UI de l'échantillon M2004.



La figure suivante met au jour par exemple la ventilation des items N et V (annotation n°3 : POS) sur la partition PERIOD constituée par les 4 premières périodes de l'UI n°1 de l'échantillon M2004 :



Enfin, la figure suivante met au jour la ventilation (en fréquence relative) de la valeur « lone » de l'annotation n°42 (package) sur la partition PLANNING_TYPE :



5. Explorer les relations de dépendance

Les différentes fonctionnalités disponibles dans le *Trameur* pour travailler avec les annotations de relations sont décrites dans la documentation du *Trameur*. On les illustre ci-dessous sur les données de la base *Rhapsodie2Trameur*.

5.1 Recherche de dépendance sur l'ensemble de la base (avec filtrage sur les items en relation)

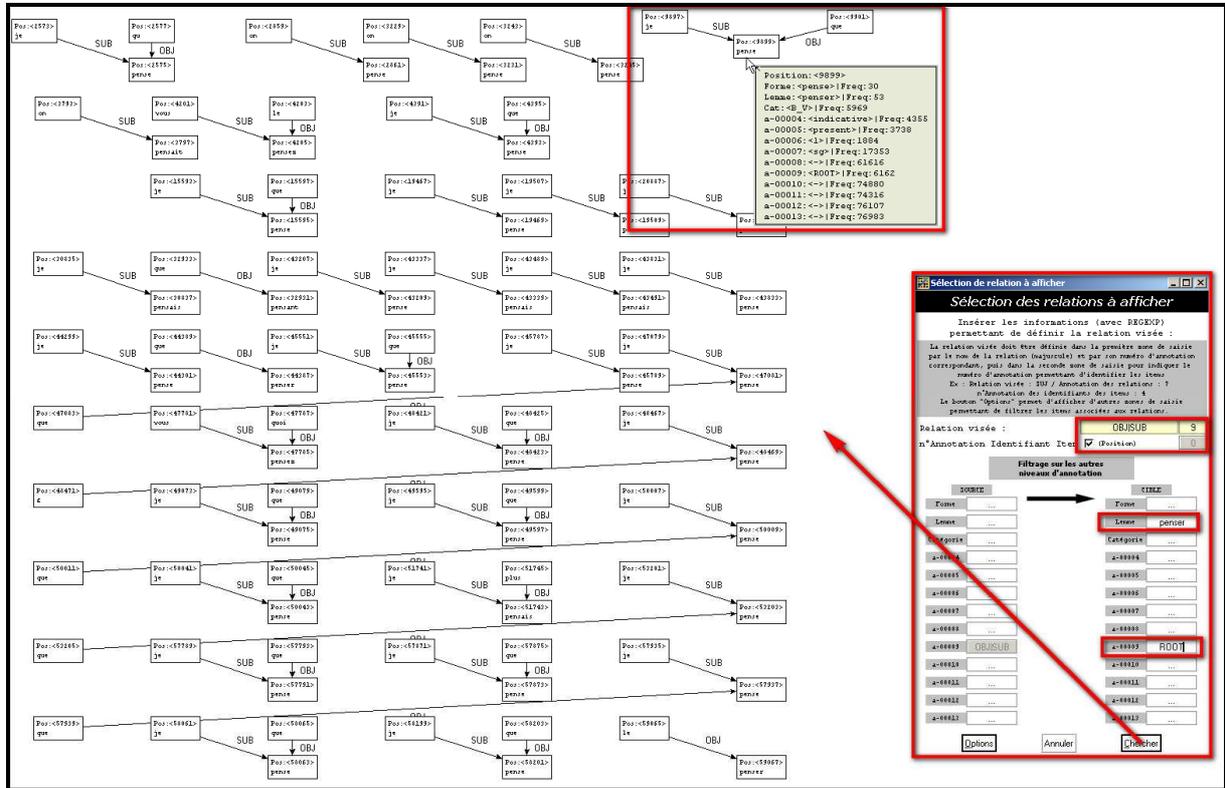
Exemple n°1 : recherche des « objets » du lemme « affirmer »

The screenshot shows the Trameur software interface. At the top, there's a menu bar with options like 'Cadre', 'Ventilation', 'Section', 'Forme-Lemme', 'Catégorie-Tag', 'Segment', 'Cocq', 'Stat', 'Concordance', 'Patron', 'Graphe', 'Sélection', 'Rapport', and 'Param'. Below the menu, there's a toolbar with icons for various functions. The main window displays a dependency graph with nodes and edges. A dialog box titled 'Sélection de relation à afficher' is open, showing a search interface. The dialog has a title bar and a main area with instructions: 'Insérer les informations (avec REGEXP) permettant de définir la relation visée :'. Below this, there's a table with columns 'SOURCE' and 'CIBLE'. The 'CIBLE' column has a row with 'Lemme' and 'affirmer'. At the bottom of the dialog, there are buttons for 'Options', 'Annuler', and 'Chercher'. Red arrows point from the search dialog to the graph and from the 'Chercher' button to the search dialog.

Dans la figure précédente, on cherche des relations de dépendance de type OBJ en contraignant un des composants de la relation : il doit porter le lemme "affirmer" (i.e. on cherche les objets d'affirmer).

Exemple n°2 : recherche des « sujets » et « objets » du lemme « penser »

Dans la figure suivante, la relation cherchée est double via l'expression régulière SUB|OBJ *i.e* SUB ou OBJ, la cible de la relation impose une valeur pour le lemme (« penser ») et pour l'annotation n°9¹ (ROOT)



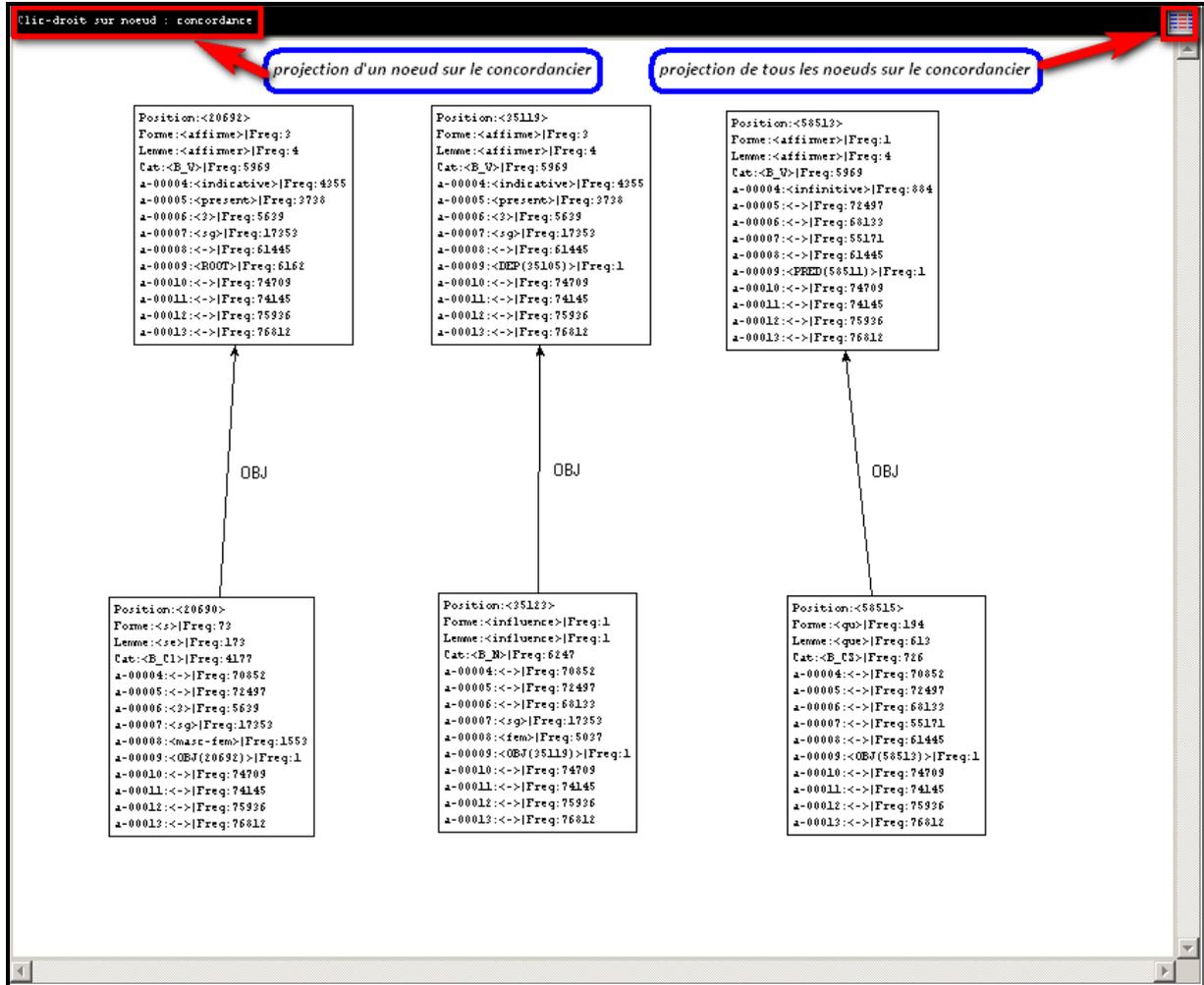
Dans les 2 graphes précédents, l’affichage des nœuds est celui produit par défaut : affichage de la position de l’item sur la *Trame* et de sa forme graphique. On peut visualiser l’ensemble des annotations d’un nœud donné via le mode Aperçu disponible dans tous les éditeurs du *Trameur* (en passant la souris au-dessus de l’objet visé) : ci-dessus, l’item en position 9899 (lemme : penser) est mis au jour.

Dans la figure qui suit, l’affichage d’un nœud est directement rendu « complet » via l’option du même nom disponible et activable avant de charger un graphe.

¹ Dans la base v7, cette valeur est portée par l’annotation n°10.

5.2 Retour en contexte

Un graphe de relations de dépendance étant produit, chaque nœud du graphe peut-être projeté sur le concordancier (clic-droit sur le nœud). L'ensemble des nœuds peut aussi être projeté globalement sur le concordancier :



La concordance produite dans ce dernier cas a l'allure suivante :

```

-----PARTIE{PARTIE=M2004}-----
      § un nouvel ordre international s'affirme peu à peu
-----PARTIE{PARTIE=M2001}-----
      qu'une nation affirme son influence qu'elle pèse dans
-----PARTIE{PARTIE=D2013}-----
      Rachida Dati pour oser affirmer qu'il ne s'agit
    
```

Les sources de la relation sont coloriées en rouge (pôle de la concordance), les cibles en vert (*i.e* ici le verbe est en vert, son objet en rouge).

On peut aussi varier l'affichage des annotations dans la concordance en matérialisant par exemple la catégorie de chaque item :

```

-----PARTIE{PARTIE=M2004}-----
$ un-B_D nouvel-B_Adj ordre-B_N international-B_Adj s-B_Cl '-I_Cl affirme-B_V peu-B_Adv à-I_Adv peu-I_Adv
-----PARTIE{PARTIE=M2001}-----
qu-B_Qu '-I_Qu une-B_D nation-B_N affirme-B_V son-B_D influence-B_N qu-B_Qu '-I_Qu elle-B_Cl pèse-B_V dans-B_Pre
-----PARTIE{PARTIE=D2013}-----
Rachida-B_N Dati-B_N pour-B_Pre oser-B_V affirmer-B_V qu-B_CS '-I_CS il-B_Cl ne-B_Cl s-B_Cl '-I_Cl agit-B_V
    
```

Toutes les zones d'édition du *Trameur* permettent de mettre au jour les annotations de la *Trame* (cf documentation en ligne, partie « Marquage des annotations de la *Trame* »). Si on considère la figure suivante, elle présente un extrait de la concordance construite à partir du graphe ayant permis d'extraire la relation OBJ (cf graphe infra) :

```

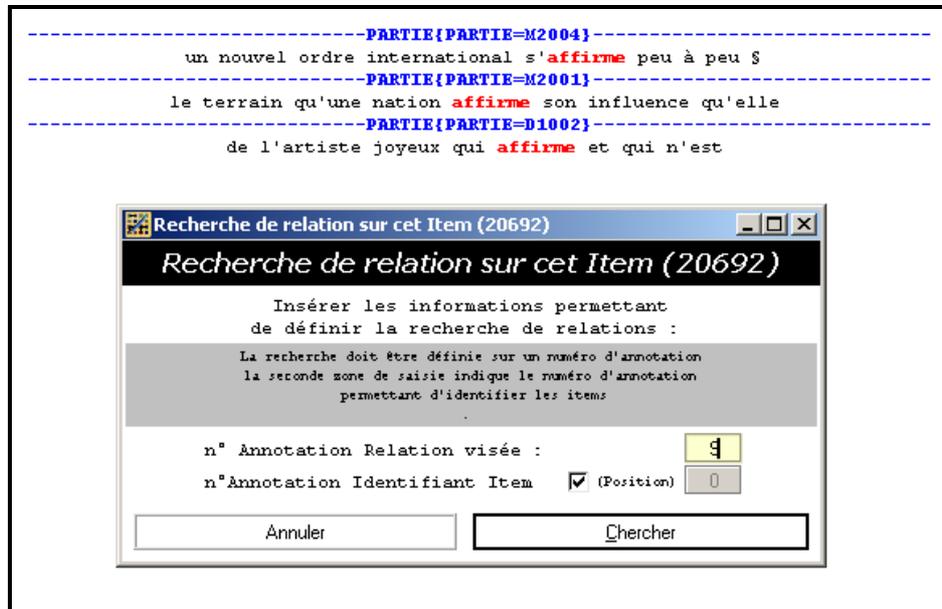
-----PARTIE{PARTIE=M2006}-----
$ de la crise aux Antilles il ne devrait pas en être question au sommet social de l'Élysée cet
savamment minütée $ alors avant d'aborder ce qui peut en sortir je vous propose de voir où quand et
alors avant d'aborder ce qui peut en sortir je vous propose de voir où quand et comment cela va
d'aborder ce qui peut en sortir je vous propose de voir où quand et comment cela va se passer
passer avec vous Jean-François Achilli $ alors passons maintenant au détail des mesures discutées et aux attentes des syndicats
des mesures de justice je cite pour les salariés touchés par la crise économique $ mais Sara Ghibaud il ne
vous d'aujourd'hui les Français se montrent très sévères à l'égard de la politique économique du gouvernement qu'
des voitures $ bonjour Sébastien Paour $ vous vous trouvez au Gosier $ où en est la situation $ merci
$ tout le monde est maintenant en tout cas suspendu à ce que Nicolas Sarkozy annoncera demain lors de sa
a promulgué son plan de relance adopté ce week-end par le congrès $ sept cent quatre-vingt-sept milliards
$ sept cent quatre-vingt-sept milliards de dollars destinés à sauver ou créer plus de trois millions et demi
à ralentir les saisies immobilières devrait être annoncé aujourd'hui par le président américain $ cela concerne plusieurs millions d'
Motors et Chrysler ont présenté hier leur plan de restructuration à l'administration Obama $ ils demandent vingt-deux milliards
toujours en garde à vue $ les enquêteurs cherchent maintenant à savoir si elle a été complice dans cette évasion
a été complice dans cette évasion $ elle se trouvait dans le parloir de la prison au moment où les
la foire d'empoigne $ la tension est encore montée d'un cran hier $ les avocats du berger corse
procès en attendant que la cour d'assise se prononce sur leur demande de supplément d'information $ lundi le
d- ce sont des procédés terroristes $ il le dit à Laurent Doulsant $ direction maintenant Barcelone pour un congrès
Laurent Doulsant $ direction maintenant Barcelone pour un congrès consacré au téléphone mobile en crise lui aussi $ pour la
nul zéro partout pour Lille au Mans $ Valenciennes sort de la zone de relégation après sa victoire sur Caen
sa victoire sur Caen deux zéro $ Nice s'impose à Nancy deux à un $
-----PARTIE{PARTIE=M0023}-----
voit euh qui voit ça passer $ et donc elle se dit eh bah elle peut récupérer elle pourra récupérer
rencontre $ et et donc elle elle tombe $ ils se & $ ouais il y a un accident quoi
ouais il y a un accident quoi $ et ils se & ils tombent tous les deux $ et c'
boulanger voit que la baguette a disparu $ et il se dit que & $ ouais moi je pense qu'
va & $ alors le euh euh le boulanger dit au policier qu'en fait c'est la fille $
    
```

Le marquage de certains items est possible en sélectionnant les annotations à mettre au jour :

The screenshot shows the Trameur software interface. On the left, there is a sidebar with various options like 'Fiche', 'Lecture', and 'Catégorie'. The main area displays a concordance table with text from the document. A dialog box titled 'Sélection des annotations à colorier...' is open, allowing the user to select specific annotations to highlight. The dialog box contains a table with columns for 'Annotation visée' and 'n°Annotation Identifiant Item (Division)'. The user has selected 'B_N' and '3_CS' in the 'Annotation visée' column and '4' in the 'n°Annotation Identifiant Item (Division)' column. A red arrow points from the dialog box to the corresponding text in the concordance table.

5.3 Recherche de dépendance en contexte

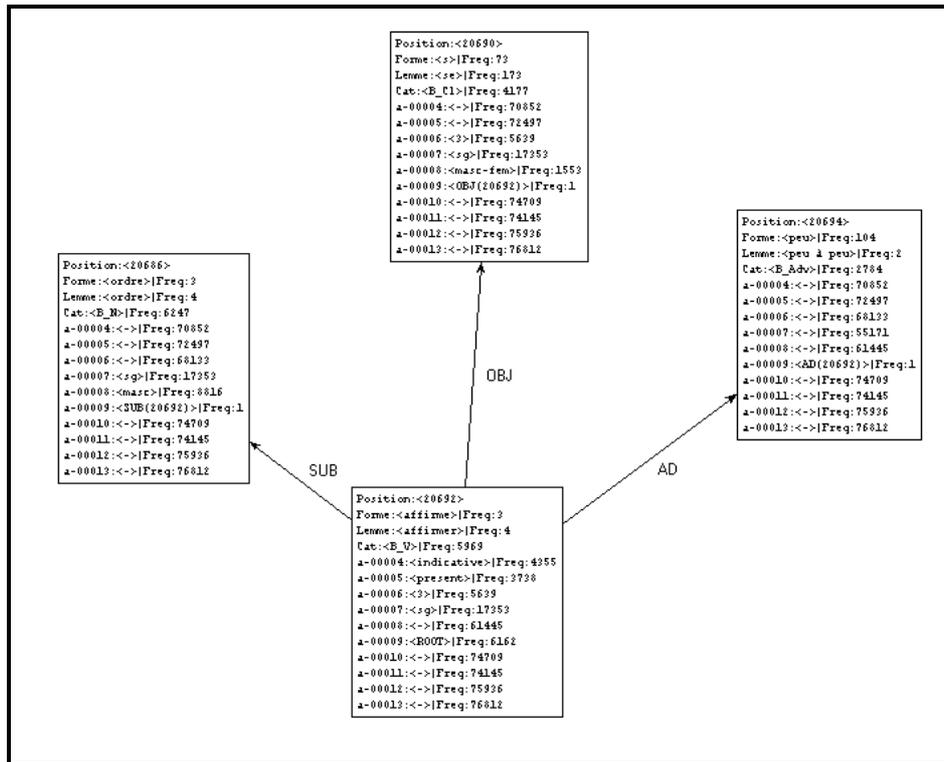
La recherche de dépendance peut aussi être réalisée en contexte, par exemple sur le concordancier. Si on considère la concordance suivante (en haut de la figure) :



Le raccourci clavier **Ctrl-Clic-droit** sur un item de la concordance (ici l'item visé est la première occurrence de la forme graphique « affirme ») permet de rechercher toutes les relations de dépendance pointant sur cet item. On commence par indiquer où chercher les relations de dépendance (numéro d'annotation portant ce type d'information : ici le n°²) et comment sont indexer les items sur la trame (ici par leur position).

La recherche conduit à la production d'un graphe mettant au jour toutes les relations sur l'item visé :

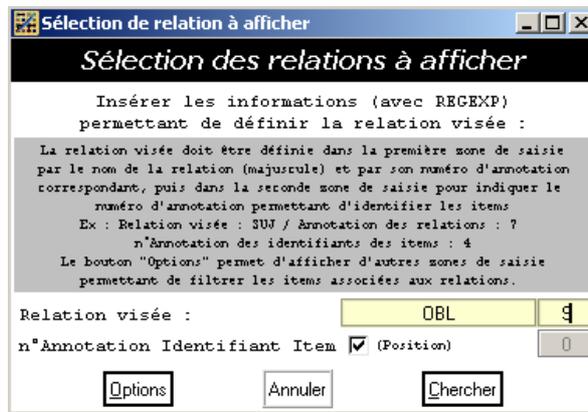
² Même remarque que dans la note précédente.



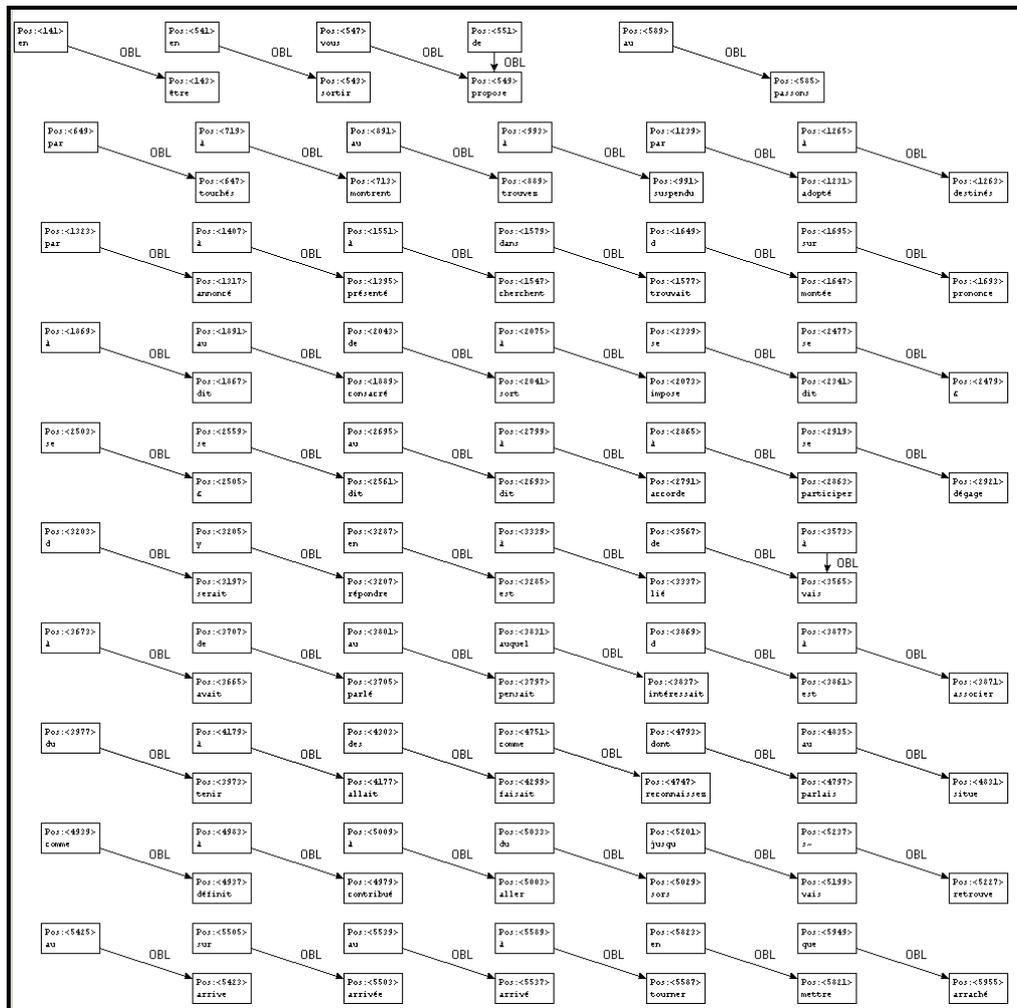
Le graphe donne à voir les 3 relations pointant sur l’item.

5.4 Rechercher dans un graphe de dépendance

On présente tout d'abord le résultat produit par la requête suivante « recherche de la relation OBL » :



Le résultat produit a l'allure suivante :



On peut mettre au jour dans ce graphe certains nœuds en filtrant leurs annotations.

Dans la figure qui suit, 3 filtrages successifs sont réalisés pour marquer successivement les nœuds ayant pour catégorie B_V, puis B_N et enfin B_Pre ; chaque requête est précédée par la modification de la couleur à utiliser pour le marquage des nœuds visés (verbe en bleu, nom en vert et préposition en orange) :



6. Recherche de collocation : spécificités sur relation

Une version particulière du module de calcul des cooccurrences permet de prendre en compte les relations entre les items de la Trame :

Etant donné une forme pôle, la recherche de ses cooccurrents peut être contrainte par la prise en compte d'une relation entre ce pôle et ses candidats cooccurrents. Pour un item X donné (le pôle), on s'intéresse aux items Y cooccurrents de X et en relation REL avec X (les collocatifs de X) :

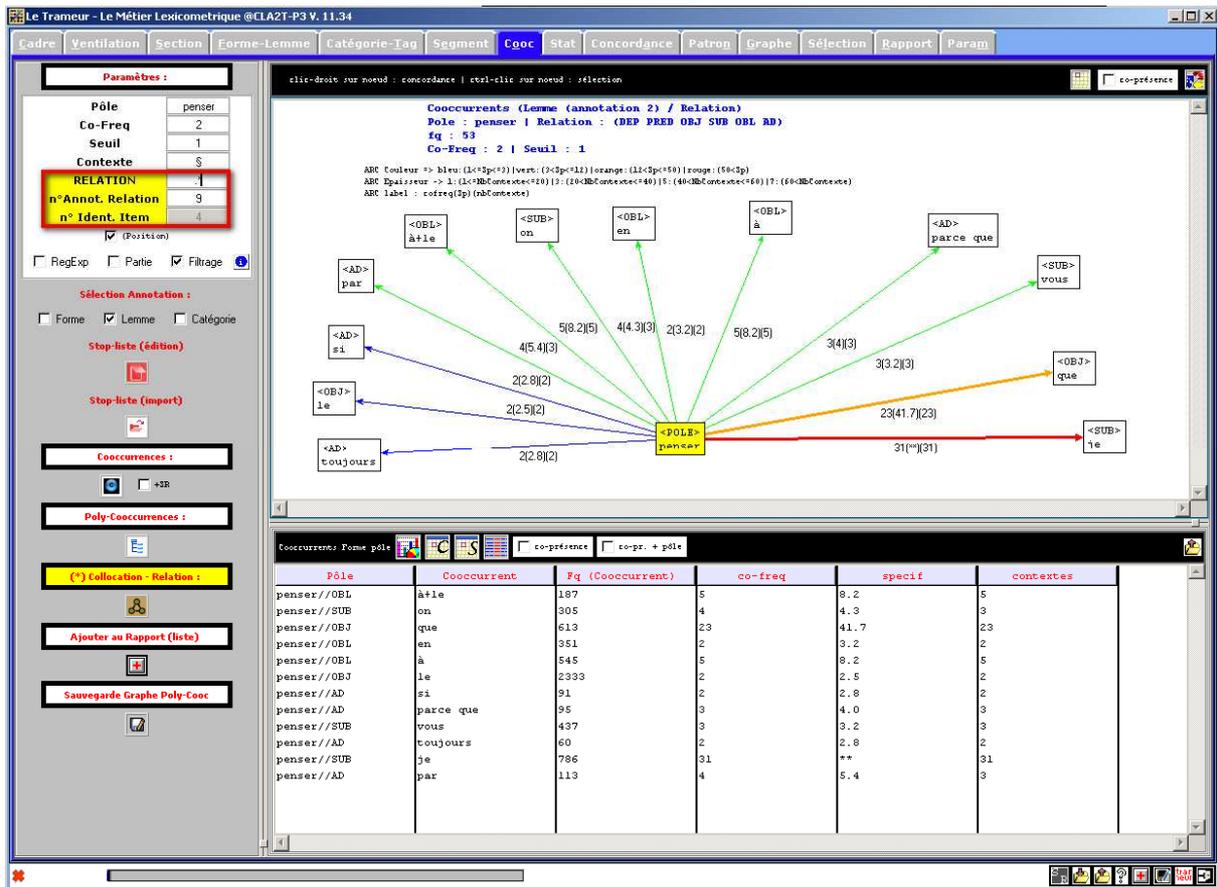
$$X \rightarrow \text{REL} \rightarrow Y$$

La figure suivante présente les cooccurrents du lemme penser en contraignant ses cooccurrents à être en position d'objet :

The screenshot shows the 'Le Trameur - Le Métier Lexicométrique @CLA2T-P3 V. 11.34' application. The 'Cooc' tab is active. The parameter panel on the left is configured with 'Pôle: penser', 'Co-Freq: 2', 'Seuil: 1', and 'Contexte: \$'. The 'RELATION' is set to 'OBJ', 'n°Annot. Relation' is 9, and 'n° Ident. Item' is 4. The 'Sélection Annotation' section has 'Lemme' checked. The main window displays a graph with a central node '<POLE> penser' connected to two nodes: '<OBJ> que' and '<OBJ> le'. The edges are labeled with '23(30.3)(23)' and '2(1.6)(2)' respectively. Below the graph is a table of cooccurrences:

Pôle	Cooccurrent	Fq (Cooccurrent)	co-freq	specif	cont
penser//OBJ	que	613	23	30.3	23
penser//OBJ	le	2333	2	1.6	2

La relation entre le pôle et ses cooccurrents peut être « non déterminée » par avance, dans la figure suivante, la relation est exprimée sous la forme : .* (regex) ; elle vise donc toutes les relations du type : penser-OBJ-y, penser-SUB-y etc.



Une fois le graphe construit, un clic-droit sur un des nœuds montrent les contextes dans lesquels la relation se réalise (*i.e* les contextes utilisés par le calcul).

Dans cet exemple, 3 contextes dans lesquels *vous* est sujet de *penser* (cf arc penser -> vous) ont été mis au jour, un clic-droit sur le nœud *vous* les montrent dans le concordancier (les 2 items de la relation y sont colorés automatiquement : *vous* en rouge (pôle de la concordance) et *penser* en vert).

```

-----PARTIE{PARTIE=D2009 }-----
un mot DELIM ou bien vous le penser vraiment DELIM
-----PARTIE{PARTIE=D0006 }-----
dans le quartier de Paris vous en penser quoi ici
-----PARTIE{PARTIE=D2011 }-----
DELIM que^êtreUNKNOWNce que vous en penser de le

```

```

Position: <64987>
Forme: <vous> | Freq: 441
Lemme: <vous> | Freq: 437
Cat: <B Cl> | Freq: 4177
a-00004: <-> | Freq: 71023
a-00005: <-> | Freq: 72668
a-00006: <2> | Freq: 988
a-00007: <p1> | Freq: 3992
a-00008: <-> | Freq: 61616
a-00009: <SUB (64991)> | Freq: 1
a-00010: <-> | Freq: 74880
a-00011: <-> | Freq: 74316
a-00012: <-> | Freq: 76107
a-00013: <-> | Freq: 76983

```